Collimation and Alignment of the Pan-STARRS PS1 Telescope

Nick Kaiser[1], Jeff Morgan[1] and Brian Stalder[2]
1) Pan-STARRS, Institute for Astronomy, University of Hawaii
2) Dept. of Physics, University of Harvard

September 4, 2009

Abstract

The Pan-STARRS PS1 telescope has 5 large optical elements and a focal plane that need to be very precisely collimated and aligned in order to deliver adequate image quality over the very large field of view. The positional tolerance is sub-millimeter. In addition, the primary mirror is supported on a bed of pneumatic supports to allow for fine control of the mirror figure. The technique we have developed to measure any mis-configuration of the elements uses linear algebraic analysis of statistics of shapes of out-of-focus images, or 'donuts'. This has proved to be remarkably successful, allowing precision of on the order of ten microns in decenter and a few arc-seconds in tips and tilts.

1 Introduction

1.1 Background to the Problem

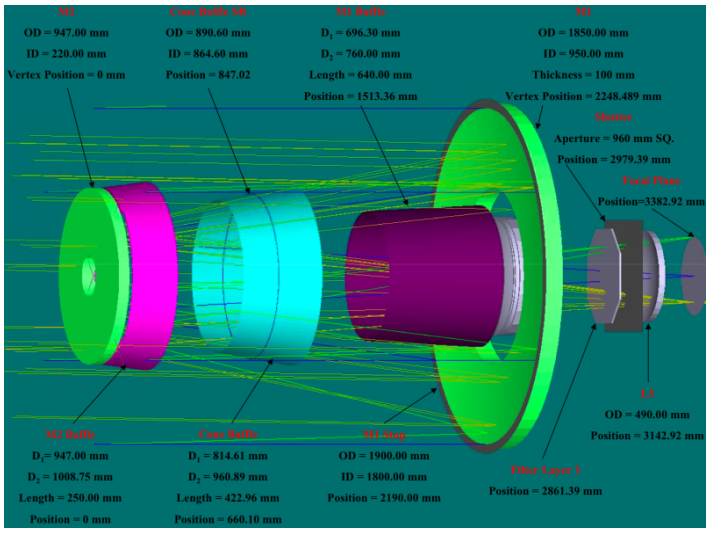
The Pan-STARRS telescope has to deliver sub-arcsecond images over a 3-degree diameter field of view. It has 2 mirrors, 3 lenses (several highly aspheric surfaces) and a focal plane as shown in figure 1. It is an alt-az design with L3 and the camera behind the instrument rotator. That gives 5x5 + 3 - 1= 27 degrees of freedom (tilts, de-centers and de-spaces) that need to be adjusted to get all of the elements collimated and aligned with the rotator axis. The required precision is sub-mm. In addition, the primary mirror support system has 12 independently controllable pneumatic axial supports, as shown in figure 1 giving a total of 39 dof under our control.

When we embarked on this problem we consulted a number of independent experts to get their advice on this problem. The results were not encouraging. Common themes were just too many parameters; that there would be problematic degeneracies; and that any iterative minimization would get stuck in local minima or valleys. Another common suggestion was to try to 'divide and conquer' by first mounting M1 and adjusting that and then adding the other elements one-by-one and adjusting their positions sequentially. While it has some attractions, this approach does not since the M1/M2 system does not deliver in focus images even on axis. A spectre hanging over the effort was the lessons learned from the CFHT MEGAPRIME, where the optical alignment had caused serious delays.

1.2 Initial Mechanical Alignment and Initial Image Quality

To get a sensible starting point, initial collimation and alignment performed with 'auto-reflecting telescope' (ART) This was mounted on a jig on instrument support stucture and aligned with the rotator axis. This was then used to align other elements using targets and fiducial marks. The upper parts of corrector optics (L1, L2) were accurately assembled as a single unit at U. Washington.

First light with GPC1 camera in July 2007 gave huge image aberrations; the images were elongated by about 10 arc-seconds in the outer part of the field. Simple modelling of the in-focus images using ray-tracing was able to give crude agreement with the observed images, and adjusting the mirrors got us down to approximately 2-arcsec IQ, still much worse than our requirement. At that point we convinced ourselves that L1 and L2 had a de-space problem - this seemed to be supported by ray-tracing - and an adjustment was made, but this actually worsened the IQ. At that point we realized we needed a more



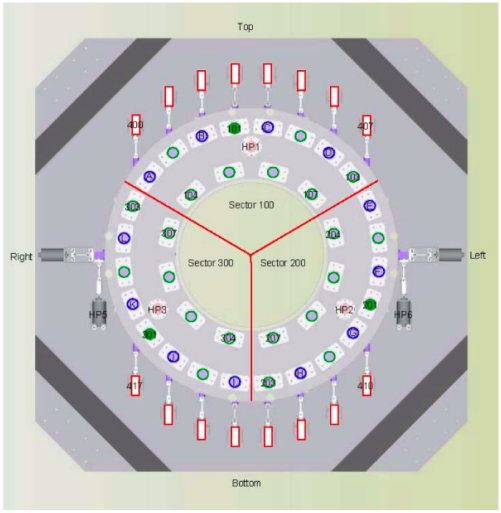


Figure 8. The M1 support system as seen from M2.

Figure 1: PS1 optical elements and PS1 primary mirror support system. M1 sits on a bed of 36 pneumatic support. These are grouped into 3 sectors which are under servo control to maintain zero load on 3 hard points in order to control the piston, tip and tilt of the mirror. In addition, 12 of the supports (colored blue in the figure) are independently adjustable to allow control of the primary figure.

methodical approach to solving the problem as analysis of in-focus images was indeed beset by problems of degeneracy.

1.3 Survey of Wave-front Diagnostic Techniques

There are many traditional (and some non-traditional) methods for diagnosing telescope aberrations. These include Shack-Hartmann imaging; Hartmann mask imaging; Knife-edge tests; Rotational beam-shearing interferometry; Ghost image analysis; Direct metrology; Direct (in-focus) image analysis; Analysis of out of focus ('extra-focal' or 'donut') images, which, in the AO community, is usually referred to as 'curvature sensing'.

We have relied primarily on full-field donut image analysis - the subject of this paper - but we have also used direct metrology, ghost images, and have been testing the built in deployable Shack-Hartmann and field-edge calcite curvature sensors.

1.4 Wavefront Aberrations from Out-of-Focus Images

1.4.1 The Basic Concept

A perfectly collimated and designed telescope gives uniformly illuminated (aside from shadowing by spider, baffles) annular out-of-focus images (the hole in the center of the annulus being caused by the hole in the primary mirror). If one is sufficiently far from focus this is well described by geometric optics - images formed by rays. Aberrations (displacement of elements or figure errors) deflect rays and cause distortion of the shape of the out-of-focus image. What is happening is crowding or dilution of the density of rays, as illustrated in figure 2 resulting in modulation of the brightness of the donuts. The local brightness is proportional to curvature (Laplacian) of the upstream wavefront curvature, hence the terminology 'curvature sensing', as pioneered at IfA by Roddier group for Adaptive Optics.[1]

[1] Curvature sensing is something of a misnomer since, in addition to the modulation of the brightness, the shape of the boundary of the donut is also distorted in a useful manner. For example, a pure astigmatism has zero curvature - the curvature in the two primary orthogonal directions being equal and opposite - so there is no modulation of the brightness, but the aberration can still be measured from out of focus images which, in this case, become elliptical.

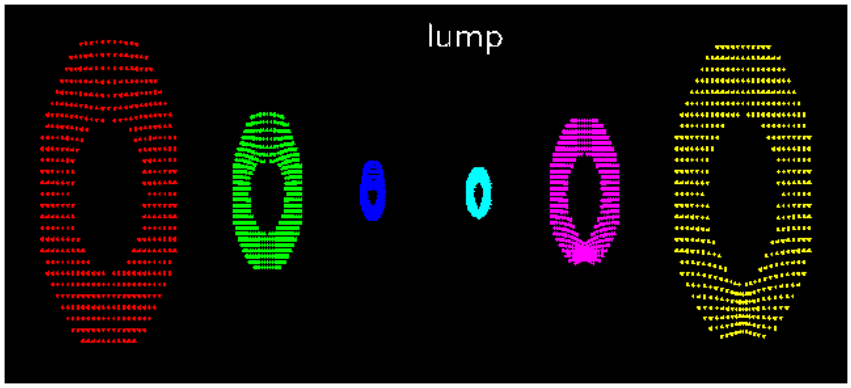


Figure 2: Illustration of the effect of a localised aberration such as might be caused by a bump on a mirror. We are plotting here the intersection of rays that span a uniform grid on the pupil plane with a series of planes displaced slightly from the true focus. On the left hand side of the figure (before focus) this has caused the rays to converge less than they should have, thus reducing the density of rays and reducing the intensity of the pre-focus donut image. Rays from this region come to focus just beyond the true focal plane, and so the result is that in a post-focus donut image there is a crowding of the rays and an increase in the donut image brightness.

1.4.2 Advantages: Linearity and Multiplexing

One advantage of donut shape analysis is that statistics derived from these are a *linear* response to the various causes of aberrations. This linearity breaks down too close to focus, but provided the donuts are much larger than the transverse displacements of the rays linearity is guaranteed. If the telescope is defocused too much then shadowing becomes problematic, and the images become very faint and this means that careful flat-fielding is needed. We have found that a 3mm defocus is a good compromise, giving good linearity and yet bright images that are easier to measure. As the optics are tuned up it is then possible to go closer to focus and still maintain linearity and this allows for more sensitivity. However, at some point this may require 'physical optics' modelling; we have not yet attempted that.

The reason that linearity is a benefit is that, barring degeneracies (more on this later), we can solve a set of linear equations to find all the displacements and needed actuator commands needed to give good IQ. All that is needed, essentially, is that one measure a sufficient number of shape statistics in order to separate the effects of the different aberrations. This gives, in principal, a 'one-step' solution — no iteration is needed so problems of failure to converge do not arise. Similarly there are, again in principle, no local minima - the method returns a unique global minumum. There are a few *caveats*; if one starts too far from the proper configuration then non-linearity may cause errors that require a further iteration. Also, it is possible to break the large d.o.f. system up into simpler sub-systems and adjust these sequentially and then, again, iteration is required.

The benefits of linearity are shared by several other techniques mentioned above (e.g. Shack-Hartmann), but an advantage notable avantage of full-field donut imaging is massive multiplexing as with PS1 one obtains thousands of donuts per exposure, and so there is a great deal of information.

1.4.3 Symmetries: Separating Seeing, Mirror Wobbles and Absorption from Aberrations

We would like to be able to measure aberrations that give ray deflections small compared to the seeing disk. Seeing causes wavefront errors of many radians of phase, and causes donuts to wobble around like jello. This averages out in long exposure, but leads to smearing. Worse still, mirror oscillations may produce persistent anisotropies, and one would not want to attribute these to collimation and alignment problems. Another source of donut shape anisotropies is shadowing by e.g. the M2 support and potentially by patchy dust on the elements.

Fortunately, such effects can be readily distinguished by using pre- and post-focus image pairs. This is illustrated in figure 3.

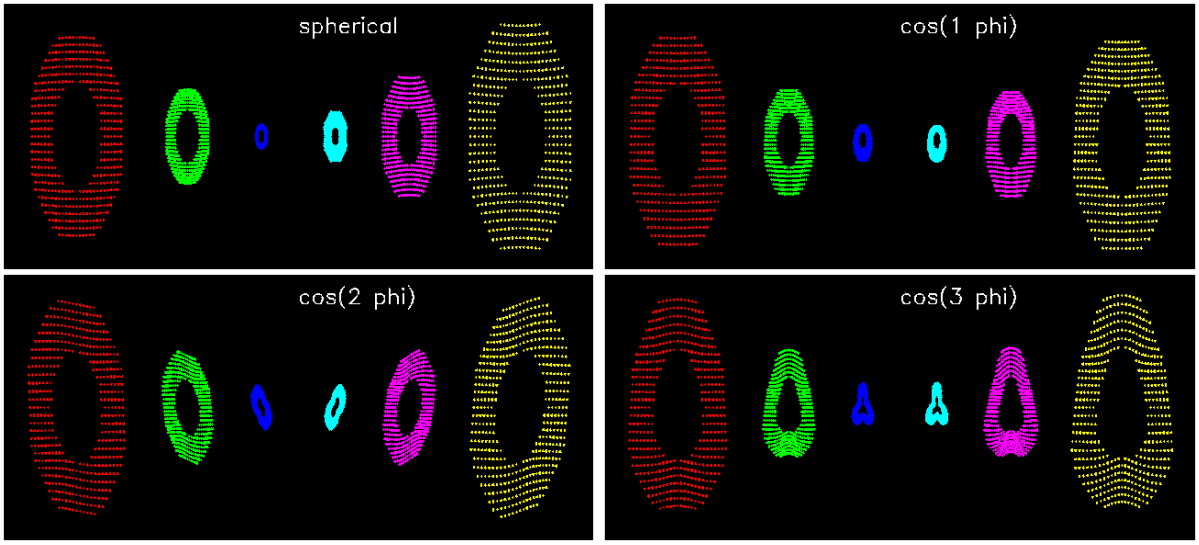


Figure 3: Symmetries of various modes of aberrations. Even angular harmonics always change sign passing through focus, while the odd harmonics do not If we rotate the post focus images by 180 degrees the sign always changes. The behaviour is opposite in the case for shadowing by masks, mirror supports or attenuation.

1.5 Degeneracies - the dreaded 'D-word'

In the present context, 'degeneracy' is usually used to describe combinations of displacements of different elements that do not cause any measurable aberration.[2] and 'quasi-degeneracy' is used to describe situations where there are combinations that produce almost zero measurable effect. These cause the linear equation solution (inversion of matrix) to be singular or ill conditioned. An exact degeneracy arises because IQ only depends on relative positioning of optical elements, but that is easy to deal with. Quasi-degeneracies are a little more tricky. One example is that decenters and or tilts of M1 are nearly degenerate with decenter of M2 in that they all produce donut distortions that look very similar — tilt of M2, on the other hand, produces a very different pattern. Similar quasi-degeneracies exist for the L3/focal-plane system.

But quasi-degeneracies are not to be feared; they are our friends. They can be identified using elementary linear algebraic techniques. They allow one to correct for one misaligment that is difficult to cure by moving another element or combination of elements that are easier. In fact the existence of quasi-degeneracies of the kind mentioned above is fundamental to Pan-STARRS mechanical design where we have real-time control over M1 and M2 but not of e.g. the camera; the designers did not assume that the camera would suffer no flexure, rather they assumed (correctly) that such deformations can be corrected, to sufficient accuracy, by adjusting the elements that are controllable.

In what follows we will first describe the pipeline we have developed to extract reliable measurements of donut shapes. We will then describe the linear algebra and how we deal with quasi-degeneracies, and we conclude with a demonstration of the method in action.

2 Donut Finding Pipeline

The goal is to obtain, from an out of focus gigapixel camera image, a grid of donut images covering the focal plane. The pipeline we have developed for this works as follows:

The first step is to determine the size of the donuts in one OTA image. To do this we first generate a cleaned log-scaled image, an example of this is shown in figure 4. We then compute the autocorrelation

[2]This terminology is rather sloppy, more commonly in linear algebra 'degeneracy' means that a matrix has non-distinct eigenvalues. Here it turns out that *all* decenter/tilts are degenerate in the this sense. This is unfortunate.

function of this image and cross-correlate with theoretical model in log(r) space. There is a rather sharp feature in the auto-correlation function at a lag equal to the diameter of the donuts and this allows a fairly precise estimate of the size of the donuts.

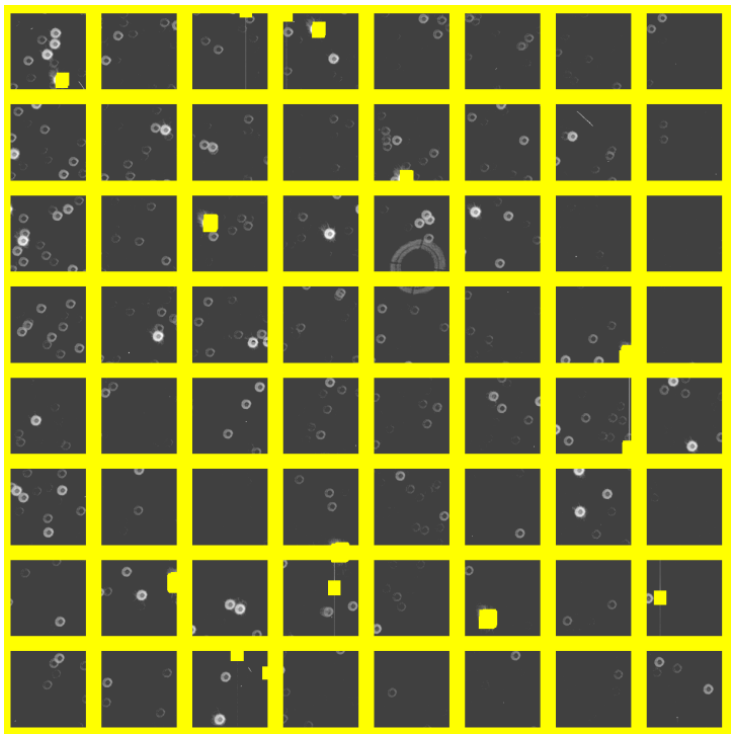


Figure 4: Log-scaled image - note the ghost.

The second step is to determine the centers of a set of high signal-to noise donuts. To do this we first generate an analytic model for the 2-D donut image, and from this construct a regularized deconvolution filter as shown in figure 5

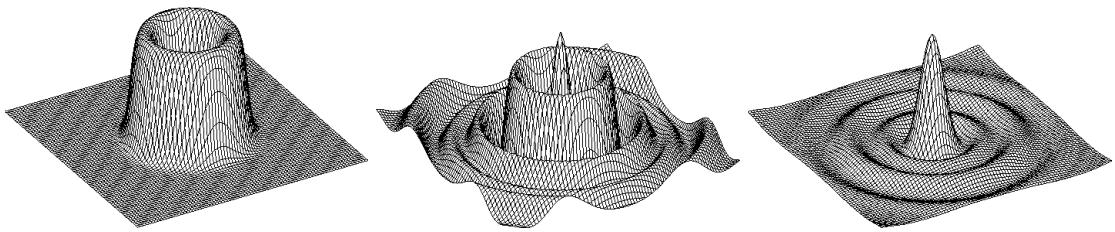


Figure 5: Donut model, regularised deconvolution filter and the convolution of the latter with the former.

Convolving this with the image and finding peaks yields the locations of a large number of candidate donuts. However, some of these have low signal to noise and others may overlap other donuts or cell boundaries. To get a clean donut image (one per detector) we first filter the catalog of candidate objects to obtain typically a few tens of good donuts per OTA and normalise these to have equal flux and then we then take median to get clean donut image. Gives a grid of 60 donuts per focal plane. The steps in this process are shown in figure 6

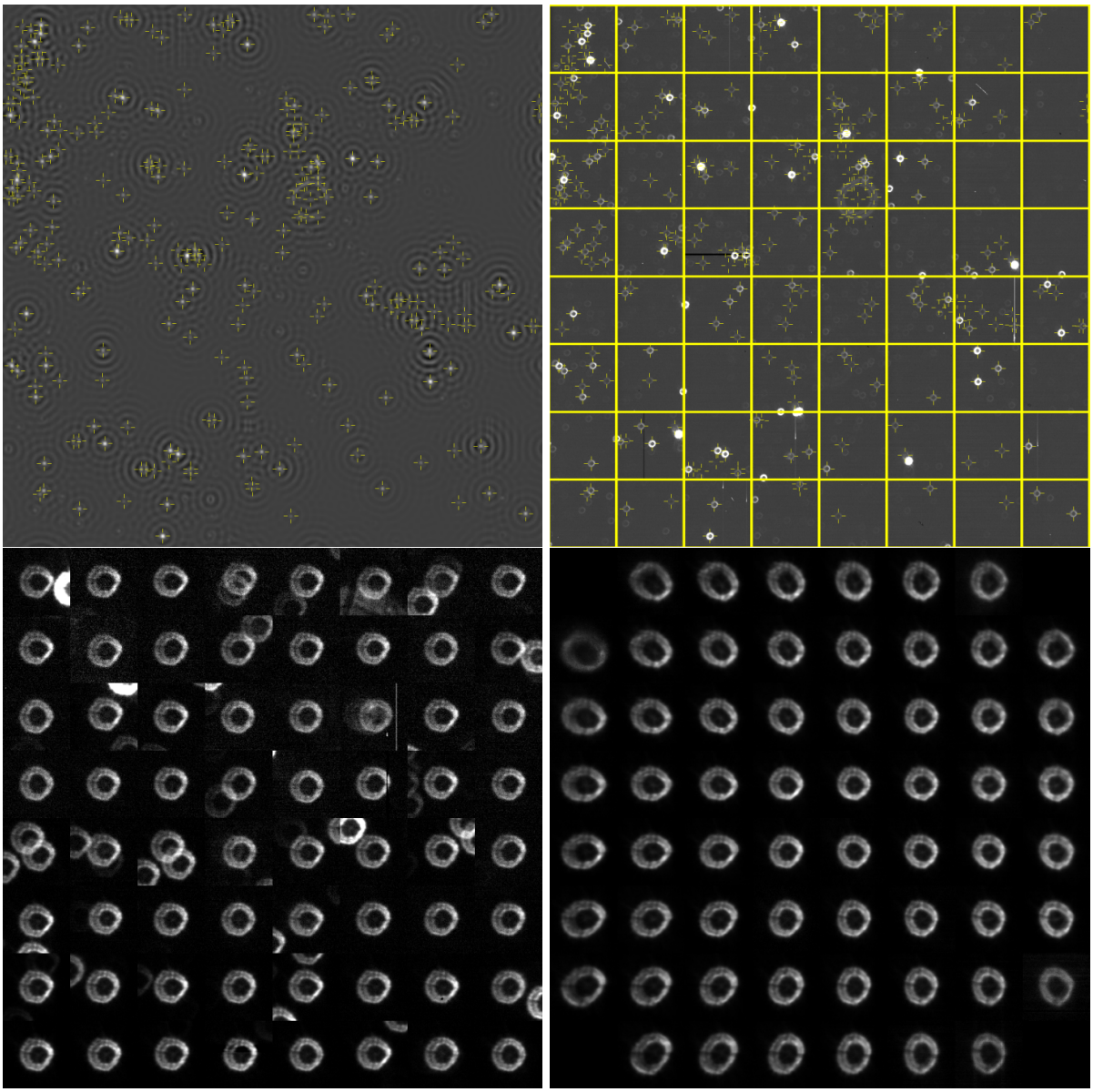


Figure 6: Upper left panel shows the log-scaled donut image convolved with the regularized deconvolution filter. The upper right panel shows the locations of the selected candidates as cross-hairs overlaid on the log-scaled image. The lower left panel shows the brighter donuts from a single OTA. Some of these are clearly corrupted, but the median of these gives a reliable single donut. The bottom right panel shows the medians for all 60 OTAs.

3 Donut Shape Statistics

It is traditional to decompose a general wave-front aberration into Zernike polynomials (for the standard terminology see Knoll, JOSA 66, p207). The first three Zernikes measure piston, tip and tilt, and do not have any effect on images, but all of the higher order modes do. Zernike mode amplitudes are effectively just moments of the wavefront (integrals of the deviation times powers of x and x) that have been made orthogonal on the unit disk, and this orthgonalisation can also be done for general pupil shapes. In the linear regime, there is a close relationship between the Zernike mode amplitudes and the analogous moments of the intensity of the donuts, with x, y measured with respect to the centroid say, so these would seem to be a natural choice of statistic, with the maximum order of the modes being a free parameter to be chosen based on the kind of aberrations that are actually generated and perfomance and rubustness considerations. The statistics that we will use are very similar to a Zernike polynomial expansion:

- We first compute the centroid of the light in a postage stamp containing the donut.
- From the pixel locations, relative to the centroid, we define a radius r and azimuthal angle.
- We then compute the mean radius in (typically 32) azimuthal bins as $r(\phi) = \sum fr / \sum f$ where f is the pixel intensity.
- For the width we typically use $w(\phi) = (\sum f)^2 / \sum f^2$. We have also tried using, as an alternative, $w(\phi) = (\sum fr^2 / \sum f - (\sum fr / \sum f))^{1/2}$but we found this to be less robust.
- We then compute Fourier coefficients of the radius, $r_l = n^{-1} \sum r(\phi) \exp(il\phi)$, the width $w_l = n^{-1} \sum w(\phi) \exp(il\phi)$ and similarly for the mean brightness, all up to some order $l_{\max}$, (typically 2, 3 or 4).
- In addition, we use the variation of the mean (zeroth order harmonic) radius and width (i.e. we subtract the average of these taken over the field in order to obtain statistics that are independent of defocus distance.
- We typically compute these statistics for each of a set of 60 positions across the focal plane; the OTA centres.

Typical example results are shown in figure 7. It is important to note that the shapes drawn are not contours of the brightness. Rather they are lines of $r(\phi)$ and $r(\phi) \pm w(\phi)/2$ generated by Fourier synthesis. Though of course the fact that they look a lot like the shape of the boundaries of the donuts is of course encouraging. The point to take away from this is that the simple, and moreover highly robust, statistics we have chosen do seem to adequately characterise the kind of donut deformation that miscollimation and misalignment actually generates.

4 Linear Algebraic Techniques for Dealing With Degeneracies

4.1 The Problem

Let's say we have a big vector of observations **s** (e.g. donut shape statistics for a bunch of stars in the present application) that is assumed to be a linear response to a smaller set of telescope element configuration parameters and/or mirror support actuators **a**

$$\begin{bmatrix} \\ \mathbf{s} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{S} \\ \\ \end{bmatrix} \cdot \begin{bmatrix} \\ \mathbf{a} \\ \\ \end{bmatrix} \tag{1}$$

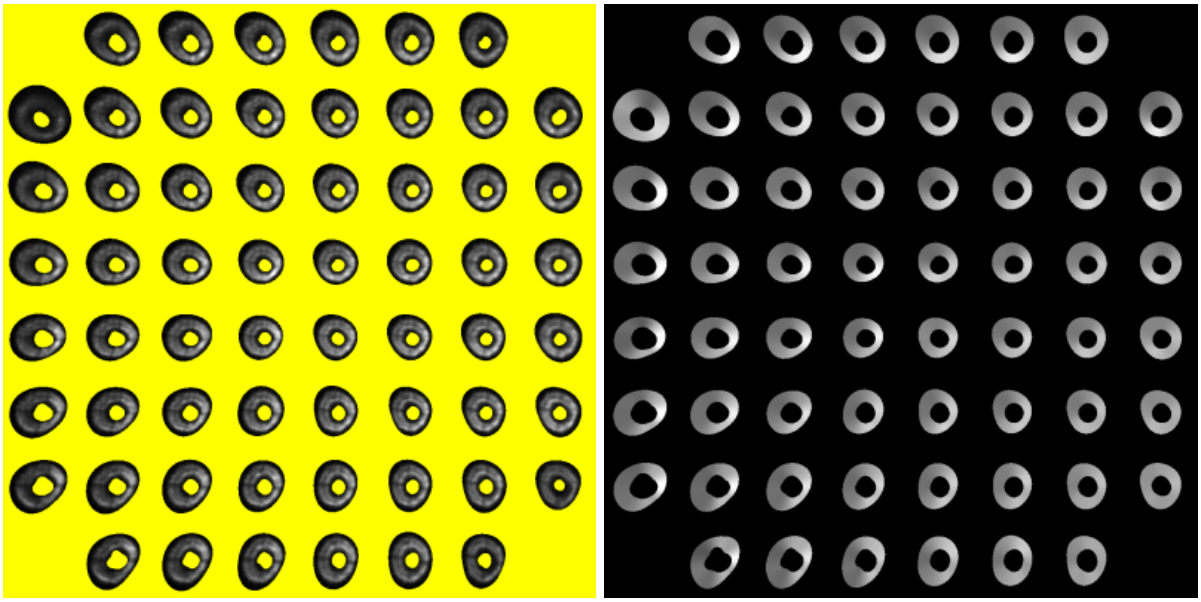


Figure 7: Operation of getdonutstats program. Masked real donuts are shown on the left. On the right is shown visualisation of the model statistics — these are donuts synthesised from the harmonic moments of the radius, width and brightness statistics.

where $\mathbf{S}$ is the 'system matrix'. We want to find a solution

$$\begin{bmatrix} \\ \mathbf{a} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{R} \\ \\ \end{bmatrix} \cdot \begin{bmatrix} \\ \mathbf{s} \\ \\ \end{bmatrix} \tag{2}$$

where $\mathbf{R}$ is the 'reconstruction matrix'. The naive approach would be to use the pseudo-inverse of the system matrix, but for a degenerate or quasi-degenerate system this will blow up in your face.

4.2 The Truncated SVD Approach

One approach is to use 'singular value decomposition' to decompose the system matrix as

$$\begin{bmatrix} \\ \mathbf{S} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{U} \\ \\ \end{bmatrix} \cdot \begin{bmatrix} \\ \mathbf{S}' \\ \\ \end{bmatrix} \cdot \begin{bmatrix} \\ \mathbf{V}^T \\ \\ \end{bmatrix} \tag{3}$$

where $\mathbf{U}$and $\mathbf{V}$ are unitary matrices (i.e. like rotation matrices) chosen to make $\mathbf{S}'$ diagonal. The idea is that the quasi-ignorable linear combinations of configuration parameters will have very small diagonal components in $\mathbf{S}'$. So if you build a reconstruction matrix $\mathbf{R} = \mathbf{V} \cdot \mathbf{S}'^{-1} \cdot \mathbf{U}^T$ but with the very large (or infinite) components of $\mathbf{S}'^{-1}$ set to zero this will give a well behaved solution. The idea is not that the solution will be close to the true solution — it won't be! But one might hope that the difference from the true solution will be ignorable (i.e. have no effect on image quality). This is neat, and widely used

in AO, but has two shortcomings. The first is that it does not 'know' about measurement noise, or noise correlations. The second, which is not widely appreciated, is that the diagonal values are not 'measure invariant'. Thus if you change the units for the configuration parameters it changes the result. This is particularly problematic here since we have inhomogeneous quantities like pressures, decenters and tilts.

4.3 The Maximum Likelihood Method

If we include additive measurement noise in the observations and say that

$$\mathbf{s} = \mathbf{S} \cdot \mathbf{a} + \mathbf{e} \tag{4}$$

and assume the errors are Gaussian distributed, with covariance matrix:

$$N_{\alpha\beta} = \langle e_\alpha e_\beta \rangle \tag{5}$$

then the likelihood is

$$L(\mathbf{a}) = P(\mathbf{s}|\mathbf{a}) = \frac{1}{\sqrt{(2\pi)^m |N|}} \exp\left\{ -\frac{1}{2} (\mathbf{s} - \mathbf{S} \cdot \mathbf{a}) \cdot \mathbf{N}^{-1} \cdot (\mathbf{s} - \mathbf{S} \cdot \mathbf{a}) \right\} \tag{6}$$

This function succinctly encapsulates all of the information at our disposal about $\mathbf{a}$ by virtue of Bayes theorem. Maximizing it gives a (relatively compact) set of linear equations

$$\begin{bmatrix} \quad \\ \mathbf{M} \\ \quad \end{bmatrix} \cdot \begin{bmatrix} \\ \mathbf{a} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \mathbf{b} \\ \\ \end{bmatrix} \tag{7}$$

where

$$\mathbf{M} = \mathbf{S}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{S} \tag{8}$$

is the 'Fisher matrix' and

$$\mathbf{b} = \mathbf{S}^T \cdot \mathbf{N}^{-1} \cdot \mathbf{s}. \tag{9}$$

This properly incorporates measurement noise and correlations thereof. Note that there is no need to try to invert the big rectangular matrix $\mathbf{S}$ it gets contracted with the noise covariance in $\mathbf{M}$ and dotted with $\mathbf{N}^{-1} \cdot \mathbf{s}$ in $\mathbf{b}$ in which $\mathbf{S}^T \cdot \mathbf{N}^{-1}$provides an optimal 'weight vector'. We can now play the same trick as with SVD and diagonalise $\mathbf{M} \rightarrow \mathbf{M}'$ but now using only elementary linear algebraic techniques and generate a reconstruction matrix by zeroing out large values of the diagonalised-frame inverse.

But that still leaves the 'measure dependence' problem. How do we deal with this? The solution we have adopted is effectively to choose units for the configuration parameters and pneumatic actuator commands etc. such that the diagonal components of the Fisher matrix (in the original, non-diagonal, frame) are all equal. That makes a unit change in any single parameter is equally observable. and means that if any component of $\mathbf{M}'$is small it is because there really is an ignorable combination of parameters and not that we have chose silly units that makes some parameter appear unimportant.

4.4 Implementation

In order to carry out the program descibed above, a necessary first step is to determine the noise covariance matrix $\mathbf{N}$. This is done empirically by exploiting the fact that the shape statistics are smooth functions of field position, so subtracting a low-order polynomial fit to the spatial dependence yields noise residuals. We assume here that the noise fluctuations in different donuts are statistically independent.

The next ingredient we need is the system matrix $\mathbf{S}$. To generate this we use ray-tracing SW to generate grid of donuts for each actuator/element displacement mode. Like ZEMAX, this uses the optical prescription and solves Snell's law, but is fast and parallelized. This results in a grid of donuts for each of the fundamental displacements. The shape statistics measured from the synthetic donuts generated by the ith displacement mode provide the ith column of the system matrix. Given $\mathbf{S}$ and $\mathbf{N}$ it is then straightforward to generate the Fisher matrix $\mathbf{M}$.

We then proceed much as one would using SVD: We find the unitary matrix $\mathbf{V}$ that diagonalises the Fisher matrix: $\mathbf{M} = \mathbf{V} \cdot \mathbf{M}' \cdot \mathbf{V}^T$. Examination of the eigenvalues tells us whether we have quasi-degeneracy, and we can then generate a solution $\mathbf{a}$ for any chosen degree of truncation. In deciding how many degrees of freedom to retain two useful diagnostics are the 'bias' matrix that tells us how the parameters in the truncated solution are biased by the actual values of the other parameters and the noise covariance matrix for the solution. The idea here is to find a compromise that reduces the noise without introducing too much bias.

The ultimate result of this analysis is a truncated set of 'signal-to-noise eigenmodes' that we 'dot' the big vector of observed shape statistics with the to obtain the truncated solution for the misconfiguration of the optics.

As mentioned above, an important example of this is the M1, M2 subsystem. If we consider tilting and decentering of these elements, we have a system with 8 degrees of freedom. Visual examination of the donut patterns generated by the various displacements shows that it is very hard to distinguish between a decenter of M2 and a tilt or decenter of M1. A tilt of M2, on the other hand, generates a very different signature. This is clearly reflected in the eigenvalues of the system matrix; 4 of these are large and 4 are much smaller. Consequently, to a good approximation, one can correct for any misconfiguration of M1 and M2 by adjusting M2 alone.

An example of carrying out this program is shown in figure 8. The distortion pattern of the synthetic donuts shown on the right is very similar in broad terms to the measured ones on the left and applying the appropriate correction to M2 does a very good job of circularizing the donuts and generating sharp in-focus images. Closer inspection reveals higher order shape distortions that are not well described by any tilt or decenter. These are the result of figure errors that we correct using the pneumatic supports of M1. Here we have used a more empirical approach and compute the matrix that connects the complex amplitudes for sinusoidal forcing terms on the mirror to the sinusoidal modulation of the radius of the donuts. The result has proven to be very successful.

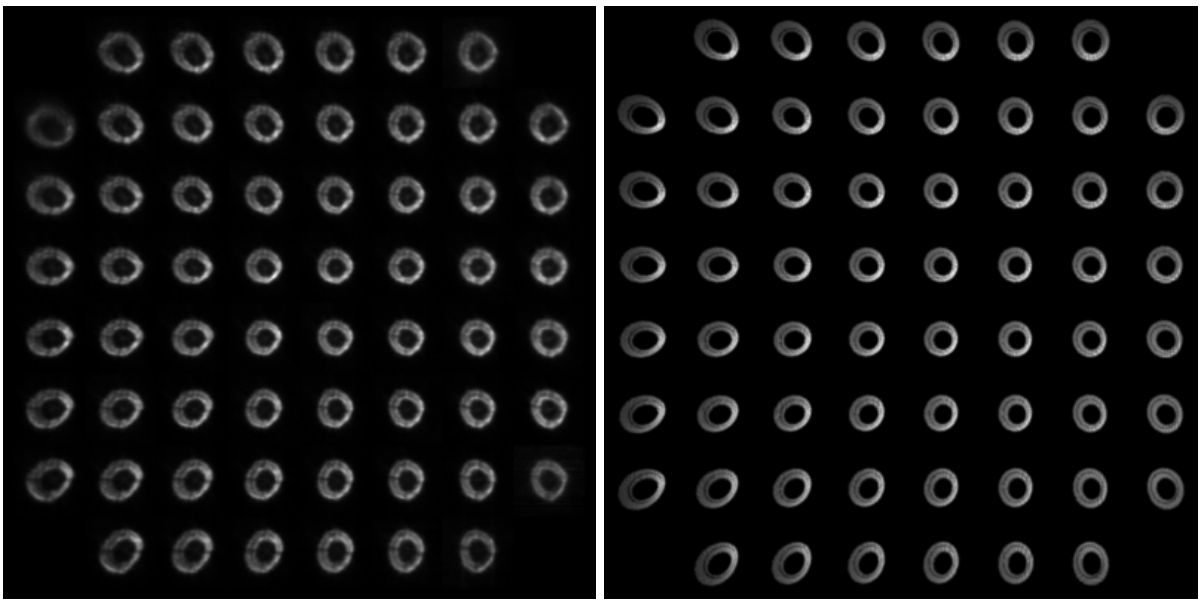


Figure 8: Donuts from exposure o4434g0351o (left) compared with ray-trace model with best fitting M1/M2 tilt and decenter.